Introduction to Focus Issue: Genetic Interactions

Introduction to Focus Issue: Genetic Interactions

Daniel Segrè[1] and Christopher J. Marx[2]
[1]*Department of Biology, Department of Biomedical Engineering, Bioinformatics Program, Boston University, Boston, Massachusetts 02215, USA*
[2]*Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, USA*

(Received 1 June 2010; published online 30 June 2010)

The perturbation of a gene in an organism's genome often causes changes in the organism's observable properties or phenotypes. It is not obvious *a priori* whether the simultaneous perturbation of two genes produces a phenotypic change that is easily predictable from the changes caused by individual perturbations. In fact, this is often not the case: the nonlinearity and interdependence between genetic variants in determining phenotypes, also known as epistasis, is a prevalent phenomenon in biological systems. This focus issue presents recent developments in the study of epistasis and genetic interactions, emphasizing the broad implications of this phenomenon in evolutionary biology, functional genomics, and human diseases. © *2010 American Institute of Physics*. [doi:10.1063/1.3456057]

A long-standing question in biology is how the instructions written in the heritable blueprint of a living system (its genotype) determine its observable properties (its phenotype).[1,2] The genotype-phenotype mapping is important for many reasons, from gaining fundamental understanding of how evolution works,[1,3] to uncovering the molecular mechanisms of genetic disease,[4,5] and identifying the "Achilles' heels" of microbial pathogens.[6,7] One way of probing this mapping is to study how different versions of the genotype (either naturally occurring or artificially generated) produce different phenotypes.[8,9] This focus issue brings together different viewpoints and approaches to understanding epistasis,[3,10–12] i.e., the nonlinearities often present in the genotype-phenotype mapping.

To introduce the concept of epistasis, let us start with a highly oversimplified view where the myriad instructions written in a genome can be thought of as discrete, binary units ("genes"), whose values determine an array of quantitative phenotypic traits. Assume for simplicity that an unperturbed organism has all genes set to zero. Single perturbation experiments may then be performed, switching individual genes to the value one. In a "first order" approximation of biology, one may be satisfied with knowing how the perturbation of each gene affects the phenotypes. For example, imagine that gene A exerts some control over metabolic rate, such that switching A from zero to one decreases the metabolic rate by 10%. It may be the case that another gene B influences metabolic rate to the same extent. The first order approximation might still work if the perturbations of multiple genes combine according to a simple, general law. For example, the metabolic rate may be affected additively, so that simultaneous switching of A or B decreases metabolic rate by 20%.

Yet, since the early days of genetics, it is known that living systems can deviate quite dramatically from such a simple linear behavior. An extreme case is one in which the effect of a double perturbation is drastically enhanced relative to the effects of individual perturbations (synergistic effect). In the example above, this may mean that the simultaneous switching of A and B gives a metabolic rate of zero, killing the organism. Or conversely, it may be the case that, even if switching A or B individually causes a 10% reduction in metabolic rate, the combined effect is still a 10% reduction. This is often called an antagonistic effect. Both the synergistic and antagonistic deviations from simple additivity constitute examples of epistasis between genes A and B. One can also say that there is an epistatic (or genetic) interaction between A and B. Note that two epistatically interacting genes do not necessarily have to interact in a physical sense. The "interaction" here means a mutual dependence in determining the phenotypic effects. Taking into account epistasis may be seen as a "second order" approach to biology, one in which effects of pairwise perturbations cannot be simply inferred from individual perturbations, but have to be explicitly evaluated.

Simplified descriptions of genotype-phenotype mapping, such as the one illustrated above, are useful for introducing the concept of epistasis. However, defining, quantifying, and understanding epistasis in real biological systems can be quite a different story, and involves a number of thorny issues. Here we do not aim at a rigorous and comprehensive review of epistasis, better left for several already existing books and reviews.[10,3,11–17] However, we wish to list some of the complications that accompany this concept. (i) First one should remember that, far from the abstraction of binary switches, epistasis was initially discovered through laborious genetics experiments, involving actual breeding of animals, plants, or microbes (see Ref. 10). (ii) While in the above example, interactions are defined between genes that are deliberately perturbed, one should think of epistasis as a pervasive phenomenon that occurs between natural variants of different genes, influencing organismal fitness and evolutionary dynamics.[18–21] (iii) Moreover, while in abstract terms, and in some laboratory experiments, it may be feasible to compare two individuals that truly differ in only one or two genes, individuals in a real population will generally differ at a large number of loci in their genomes. In fact, the study of

epistasis in populations of individuals constitutes almost a discipline on its own, with a strong statistical flavor, and major potential population genetics implications.[10,11,16] (iv) In the above example, we have assumed an additive behavior as the null hypothesis of how two genes would combine in absence of epistasis. Other null hypotheses (e.g., a multiplicative behavior) are widely employed. This is a debated issue, and the subject of active research. Defining and understanding the baseline of how one would expect perturbations to combine is a key aspect of defining deviations from such null expectation.[12] (v) Epistasis is introduced above in the case where both mutations, as well as their combination, decrease fitness; this does not have to be the case. More complex cases of epistasis can be contemplated and observed, such as sign epistasis, where a mutation could increase or decrease fitness based on the underlying genetic background.[22] (vi) Epistasis can occur between any two mutations in a genome, including within a single protein.[23,24] Intra-allelic dominance in diploid organisms could be seen as a special case of epistasis as well. (vii) Finally, although epistasis is formally defined as a genetic interaction, analogous phenomena occur between pairs of environmental perturbations (such as the combined effects of different drugs) or between genetic and environmental changes (termed genotype-by-environment interactions or $G \times E$).[25–29]

The study of epistasis has been the subject of ongoing research for more than a hundred years.[10] Recently, the advent of high throughput technologies (such as cheap DNA sequencing and sophisticated robotics) and the increasing awareness that epistasis may play a central role in understanding and fighting genetic and infectious diseases have brought this field to the forefront of multidisciplinary biological research.[14,30–38] From a functional genomics perspective, the study of epistasis is motivated by the intuition that the lack of independence entailed by an epistatic interaction conveys useful information about functional relatedness of the interacting genes. In yeast, this intuition was confirmed experimentally by high-throughput studies of synthetic lethality (an extreme case of epistasis where single genes are dispensable, while the double deletion is lethal). Notably, the availability of high-throughput experimental and computational technologies makes it possible to explore simultaneously several epistatic interactions between individual genes, giving rise to epistatic interaction networks.

This capacity to efficiently screen phenotypes for thousands of different mutant strains or chemical perturbations has produced large amounts of genetic interaction data.[39] Organizing these data and interpreting its biological meaning is a nontrivial task. Two articles in this focus issue address the problem of how to organize genetic interaction data into biologically meaningful modules. The contribution by Carter *et al.*[40] highlights the relevance of information theory to tease out the differences between genetic modules and classically defined pathways. The work by Guo *et al.*[41] introduces a new recursive maximum-likelihood approach to partition an interaction network into modules, and applies it to a gene-drug interaction data set. Both papers emphasize that research beyond classical clustering algorithms is necessary in order to understand how genetic interactions are organized, and how to gain novel biological and medical insight out of them. While experimental efforts are blooming, several aspects of epistasis are still beyond experimental reach. A notable example is the question of whether the second order biology of pairwise interactions is sufficient to understand the complexity of cellular organization. Genome scale models of metabolism are a good platform for investigating computationally large numbers of enzyme perturbations in search for patterns of epistasis.[38] Imielinski and Belta[42] in this issue use genome-scale stoichiometric models to search for sets (rather than pairs) of synthetic lethal genes in human metabolism, i.e., sets of genes whose individual effect is negligible, but which can collectively impair cell growth. By implementing a search for minimal cut sets, they map alternative routes in metabolic networks, with potential applications in the study of metabolic disease.

In addition to mapping and interpreting cell-scale genetic interaction networks, it is important to understand how individual epistatic links, or overall trends in epistasis between genes in a system, affect evolutionary adaptation.[3,18–23,43–49] The relationship between epistasis and evolution has been studied in many different contexts, ranging from the adaptive evolution of viruses to the emergence of recombination and sexual reproduction, often using formal approaches. Since evolution cannot foresee the potential synergistic or antagonistic effects on one mutation on the background of another, the interplay between rate of mutations, population parameters, and epistasis can dramatically affect evolutionary dynamics. A notable, extremely active research area along these lines is the study of antibiotic resistance of microbes. For example, it has been recently found that interactions between drugs may surprisingly reverse the evolution and rise of resistant strains.[6]

In this issue, Dawid *et al.*[50] explore the connection between epistasis and ruggedness of the fitness landscape in a classical bacterial genetic regulatory network system. In addition to mapping a multipeak fitness landscape based on mutation data, they identify epistatic effects that may play a fundamental role in dictating the adaptation dynamics on this landscape. As addressed by the work by Elena *et al.*,[51] specific patterns of epistasis can be identified in RNA viruses, with predictable evolutionary consequences. In particular, epistasis in RNA viruses seems to be dominated by antagonistic effects, possibly leading to an increase in the cost of mutations (the mutational load).

A fascinating aspect of recent research on epistasis and evolution is the possibility to perform long-term evolutionary adaptation experiments on microbial systems, in order to study evolutionary dynamics and epistasis under controlled laboratory settings, and (almost) arbitrarily tailored selection pressure.[52,53] Two examples of this type of experiment are reported in this issue. The work by Pena *et al.*[54] shows that the nonlinearity in the function relating enzyme activity to fitness is sufficient to explain at a population level the dynamics of sweeps and clonal interference seen during adaptation to a changing environment. From a different perspective, Ogbunugafor *et al.*[55] describe the concept of environmental robustness. By evolving viruses under UV light, they explore the possibility of performing selection for genetic robustness, with potential important consequences in the fight against infectious diseases.

In conclusion, the concept of epistasis captures some key aspects of the nonlinearity of living systems. A combination

of experimental, computational, and theoretical studies will continue to uncover fundamental biological principles associated with epistasis. This may happen for example through the search for a modular organization of nonlinearities measured in high throughput experiments, or through the study of the interplay between epistasis and evolutionary dynamics. One should keep in mind, though, that in the jungle of biology idiosyncrasies and details matter, and that beauty, fundamental understanding, and biomedical applications will likely be hidden in those details as well.

The authors would like to acknowledge the support from the National Institutes of Health (Grant No. 1R01 GM078209). D.S. also acknowledges funding from NIH Grant No. 1RC2GM092602-01, and NASA (NASA Astrobiology Institute, NNA08CN84A

[1]D. L. Hartl and A. G. Clark, *Principles of Population Genetics*, 3rd ed. (Sinauer Associates, Sunderland, Massachusetts, 1997).
[2]R. D. Dowell, O. Ryan, A. Jansen, D. Cheung, S. Agarwala, T. Danford, D. A. Bernstein, P. A. Rolfe, L. E. Heisler, B. Chin, C. Nislow, G. Giaever, P. C. Phillips, G. R. Fink, D. K. Gifford, and C. Boone, Science **328**, 469 (2010).
[3]J. B. Wolf, E. D. Brodie III, and M. J. Wade, *Epistasis and the Evolutionary Process*, 1st ed. (Oxford University Press, New York, 2000).
[4]R. L. Nagel, C. R. Biol. **328**, 606 (2005).
[5]S. K. Sieberts and E. E. Schadt, Mamm Genome **18**, 389 (2007).
[6]R. Chait, A. Craney, and R. Kishony, Nature (London) **446**, 668 (2007).
[7]R. T. Cirz, J. K. Chin, D. R. Andes, V. de Crécy-Lagard, W. A. Craig, and F. E. Romesberg, PLoS Biol. **3**, e176 (2005).
[8]A. M. Dudley, D. M. Janse, A. Tanay, R. Shamir, and G. M. Church, Mol. Syst. Biol. 1:0001 (2005).
[9]I. M. Ehrenreich, N. Torabi, Y. Jia, J. Kent, S. Martis, J. A. Shapiro, D. Gresham, A. A. Caudy, and L. Kruglyak, Nature (London) **464**, 1039 (2010).
[10]P. C. Phillips, Nat. Rev. Genet. **9**, 855 (2008).
[11]H. J. Cordell, Hum. Mol. Genet. **11**, 2463 (2002).
[12]R. Mani, R. P. St. Onge, J. L. Hartman, G. Giaever, and F. P. Roth, Proc. Natl. Acad. Sci. U.S.A. **105**, 3461 (2008).
[13]J. H. Moore and S. M. Williams, Am. J. Hum. Genet. **85**, 309 (2009).
[14]C. Boone, H. Bussey, and B. J. Andrews, Nat. Rev. Genet. **8**, 437 (2007).
[15]H. Shao, L. C. Burrage, D. S. Sinasac, A. E. Hill, S. R. Ernest, W. O'Brien, H. Courtland, K. J. Jepsen, A. Kirby, E. J. Kulbokas, M. J. Daly, K. W. Broman, E. S. Lander, and J. H. Nadeau, Proc. Natl. Acad. Sci. U.S.A. **105**, 19910 (2008).
[16]J. H. Moore and S. M. Williams, BioEssays **27**, 637 (2005).
[17]J. M. Pérez-Pérez, H. Candela, and J. L. Micol, Trends Genet. **25**, 368 (2009).
[18]G. Martin, S. F. Elena, and T. Lenormand, Nat. Genet. **39**, 555 (2007).
[19]A. DeLuna, K. Vetsigian, N. Shoresh, M. Hegreness, M. Colón-González, S. Chao, and R. Kishony, Nat. Genet. **40**, 676 (2008).
[20]S. F. Elena and R. E. Lenski, Nature (London) **390**, 395 (1997).
[21]J. A. G. M. de Visser and S. F. Elena, Nat. Rev. Genet. **8**, 139 (2007).
[22]D. M. Weinreich, R. A. Watson, and L. Chao, Evolution (Lawrence, Kans.) **59**, 1165 (2005).
[23]D. M. Weinreich, N. F. Delaney, M. A. Depristo, and D. L. Hartl, Science **312**, 111 (2006).
[24]S. Bershtein, M. Segal, R. Bekerman, N. Tokuriki, and D. S. Tawfik, Nature (London) **444**, 929 (2006).
[25]P. Yeh, A. I. Tschumi, and R. Kishony, Nat. Genet. **38**, 489 (2006).
[26]P. Yeh and R. Kishony, Mol. Syst. Biol. **3**, 85 (2007).
[27]J. Lehár, G. R. Zimmermann, A. S. Krueger, R. A. Molnar, J. T. Ledell, A. M. Heilbut, G. F. Short, L. C. Giusti, G. P. Nolan, O. A. Magid, M. S. Lee, A. A. Borisy, B. R. Stockwell, and C. T. Keith, Mol. Syst. Biol. **3**, 80 (2007).
[28]A. Lopez, A. B. Parsons, C. Nislow, G. Giaever, and C. Boone, Prog. Drug Res. **66**, 237 (2008).
[29]M. E. Hillenmeyer, E. Ericson, R. W. Davis, C. Nislow, D. Koller, and G. Giaever, Genome Biol. **11**, R30 (2010).
[30]S. L. Ooi, X. Pan, B. D. Peyser, P. Ye, P. B. Meluh, D. S. Yuan, R. A. Irizarry, J. S. Bader, F. A. Spencer, and J. D. Boeke, Trends Genet. **22**, 56 (2006).
[31]G. Giaever, A. M. Chu, L. Ni, C. Connelly, L. Riles, S. Véronneau, S. Dow, A. Lucau-Danila, K. Anderson, B. André, A. P. Arkin, A. Astromoff, M. El-Bakkoury, R. Bangham, R. Benito, S. Brachat, S. Campanaro, M. Curtiss, K. Davis, A. Deutschbauer, K. Entian, P. Flaherty, F. Foury, D. J. Garfinkel, M. Gerstein, D. Gotte, U. Güldener, J. H. Hegemann, S. Hempel, Z. Herman, D. F. Jaramillo, D. E. Kelly, S. L. Kelly, P. Kötter, D. LaBonte, D. C. Lamb, N. Lan, H. Liang, H. Liao, L. Liu, C. Luo, M. Lussier, R. Mao, P. Menard, S. L. Ooi, J. L. Revuelta, C. J. Roberts, M. Rose, P. Ross-Macdonald, B. Scherens, G. Schimmack, B. Shafer, D. D. Shoemaker, S. Sookhai-Mahadeo, R. K. Storms, J. N. Strathern, G. Valle, M. Voet, G. Volckaert, C. Wang, T. R. Ward, J. Wilhelmy, E. A. Winzeler, Y. Yang, G. Yen, E. Youngman, K. Yu, H. Bussey, J. D. Boeke, M. Snyder, P. Philippsen, R. W. Davis, and M. Johnston, Nature (London) **418**, 387 (2002).
[32]X. Pan, D. S. Yuan, S. Ooi, X. Wang, S. Sookhai-Mahadeo, P. Meluh, and J. D. Boeke, Methods **41**, 206 (2007).
[33]D. K. Breslow, D. M. Cameron, S. R. Collins, M. Schuldiner, J. Stewart-Ornstein, H. W. Newman, S. Braun, H. D. Madhani, N. J. Krogan, and J. S. Weissman, Nat. Methods **5**, 711 (2008).
[34]S. R. Collins, M. Schuldiner, N. J. Krogan, and J. S. Weissman, Genome Biol. **7**, R63 (2006).
[35]M. Schuldiner, S. R. Collins, N. J. Thompson, V. Denic, A. Bhamidipati, T. Punna, J. Ihmels, B. Andrews, C. Boone, J. F. Greenblatt, J. S. Weissman, and N. J. Krogan, Cell **123**, 507 (2005).
[36]A. H. Y. Tong, G. Lesage, G. D. Bader, H. Ding, H. Xu, X. Xin, J. Young, G. F. Berriz, R. L. Brost, M. Chang, Y. Chen, X. Cheng, G. Chua, H. Friesen, D. S. Goldberg, J. Haynes, C. Humphries, G. He, S. Hussein, L. Ke, N. Krogan, Z. Li, J. N. Levinson, H. Lu, P. Ménard, C. Munyana, A. B. Parsons, O. Ryan, R. Tonikian, T. Roberts, A. Sdicu, J. Shapiro, B. Sheikh, B. Suter, S. L. Wong, L. V. Zhang, H. Zhu, C. G. Burd, S. Munro, C. Sander, J. Rine, J. Greenblatt, M. Peter, A. Bretscher, G. Bell, F. P. Roth, G. W. Brown, B. Andrews, H. Bussey, and C. Boone, Science **303**, 808 (2004).
[37]M. Costanzo, A. Baryshnikova, J. Bellay, Y. Kim, E. D. Spear, C. S. Sevier, H. Ding, J. L. Y. Koh, K. Toufighi, S. Mostafavi, J. Prinz, R. P. St. Onge, B. VanderSluis, T. Makhnevych, F. J. Vizeacoumar, S. Alizadeh, S. Bahr, R. L. Brost, Y. Chen, M. Cokol, R. Deshpande, Z. Li, Z. Lin, W. Liang, M. Marback, J. Paw, B. San Luis, E. Shuteriqi, A. H. Y. Tong, N. van Dyk, I. M. Wallace, J. A. Whitney, M. T. Weirauch, G. Zhong, H. Zhu, W. A. Houry, M. Brudno, S. Ragibizadeh, B. Papp, C. Pál, F. P. Roth, G. Giaever, C. Nislow, O. G. Troyanskaya, H. Bussey, G. D. Bader, A. Gingras, Q. D. Morris, P. M. Kim, C. A. Kaiser, C. L. Myers, B. J. Andrews, and C. Boone, Science **327**, 425 (2010).
[38]D. Segrè, A. Deluna, G. M. Church, and R. Kishony, Nat. Genet. **37**, 77 (2005).
[39]J. L. Y. Koh, H. Ding, M. Costanzo, A. Baryshnikova, K. Toufighi, G. D. Bader, C. L. Myers, B. J. Andrews, and C. Boone, Nucleic Acids Res. **38**, D502 (2010).
[40]G. W. Carter, C. G. Rush, F. Uygun, N. A. Sakhanenko, D. J. Galas, and T. Galitski, Chaos **20**, 026102 (2010).
[41]J. Guo, D. Tian, B. A. McKinney, and J. L. I. Hartman, Chaos **20**, 026103 (2010).
[42]M. Imielinski and C. Belta, Chaos **20**, 026104 (2010).
[43]S. Kryazhimskiy, G. Tkacik, and J. B. Plotkin, Proc. Natl. Acad. Sci. U.S.A. **106**, 18638 (2009).
[44]N. Price and I. Shmulevich, Curr. Opin. Biotechnol. **18**, 365 (2007).
[45]L. Jasnos and R. Korona, Nat. Genet. **39**, 550 (2007).
[46]T. F. Cooper, S. K. Remold, R. E. Lenski, and D. Schneider, PLoS Genet. **4**:e35 (2008).
[47]P. Gros, H. Le Nagard, and O. Tenaillon, Genetics **182**, 277 (2009).
[48]C. Adami, Nat. Rev. Genet. **7**, 109 (2006).
[49]S. S. Chow, C. O. Wilke, C. Ofria, R. E. Lenski, and C. Adami, Science **305**, 84 (2004).
[50]A. Dawid, D. Kiviet, M. Kogenaru, and S. Tans, Chaos **20**, 026105 (2010).
[51]S. F. Elena, R. V. Sole, and J. Sardanyes, Chaos **20**, 026106 (2010).
[52]H. Chou, J. Berthet, and C. J. Marx, PLoS Genet. **5**:e1000652 (2009).
[53]J. E. Barrick, D. S. Yu, S. H. Yoon, H. Jeong, T. K. Oh, D. Schneider, R. E. Lenski, and J. F. Kim, Nature (London) **461**, 1243 (2009).
[54]M. I. Pena, E. V. Itallie, M. R. Bennett, and Y. Shamoo, Chaos **20**, 026107 (2010).
[55]C. B. Ogbunugafor, J. B. Pease, and P. E. Turner, Chaos **20**, 026108 (2010).